Title: Efficient and Robust Classification of Seismic Data using Nonlinear Support Vector Machines

Author(s): Hickmann, Kyle Scott
Hyman, Jeffrey De'Haven
Srinivasan, Gowri

Intended for: Asilomar Conference on Signals, Systems, and Computers, 2017-10-29 (Asilomar, California, United States)

Issued: 2017-11-20

Efficient and Robust Classification of Seismic Data using Nonlinear Support Vector Machines

Kyle S. Hickmann
X-Computational Physics Division
Los Alamos National Laboratory
Los Alamos, NM 87545, USA
Jeffrey Hyman
Earth and Environmental
Sciences Division
Los Alamos National Laboratory
Los Alamos, NM 87545, USA
Gowri Srinivasan
Theoretical Division
Los Alamos National Laboratory
Los Alamos, NM 87545, USA
Email:gowri@lanl.gov

Abstract*—We characterize the robustness and scalability of nonlinear Support Vector Machines (SVM) combined with kernel Principal Component Analysis (kPCA) for the classification of nonlinearly correlated data within the context of geo-structure identification using seismic data. Classification through pattern recognition using supervised learning algorithms such as SVM is popular in many fields. However, the suitability of such methods for classifying seismic data is severely hampered by assumptions of linearity (linear SVM), which affects accuracy, or computational limitations with increases in data dimension (nonlinear SVM). We propose an alternate approach to overcome this limitation, performing nonlinear SVM in a reduced dimensional space determined using kPCA. We quantify trade off between accuracy and required computational time of this approach to classify nonlinearly correlated seismic data. The utility of the method is demonstrated by characterizing the geologic structure using synthetically generated seismograms. We observe that our method produced a more efficient and robust classifier for seismic data than standard nonlinear SVM. In most cases considered, optimal SVM performance occurs when a subspace that makes up only 10% of the entire feature space is used for the training set. We also observe a greater than five times speedup in computational time between the optimal performance and standard nonlinear SVM. The results indicate that performing kPCA dimension reduction prior to SVM classification can significantly increase performance, reliability, and robustness of the classifier in seismic problems.

I. INTRODUCTION

Seismic data is used by geologists for a variety of applications including earthquake monitoring, detection and classification of man-made seismic disturbances, and characterization of structures in the earth's subsurface. These applications play a key role in many defense and energy applications [1], [2]. For example, a central theme in nuclear non-proliferation is to locate and characterize seismic events to distinguish natural events, e.g., earthquakes, from anthropogenic ones, e.g., underground nuclear weapons testing [3], [4], [5]. The recent claim by North Korea regarding the testing of a hydrogen bomb is one such instance where seismic signatures were analyzed to determine the nature of the event. In the energy sector, identifying the possibility of induced seismic activity, both minor and major earthquakes, resulting from the disposal of waste water used in hydraulic fracturing operations is a pivotal concern [6], [7]. In applications such as these, identifying and classifying the source and strength of the event accurately is of utmost importance.

In recent years, seismic imaging techniques have advanced significantly, leading to higher fidelity data being recorded. Using thousands of terabytes of seismogram data to estimate a large number of nonlinearly related parameters in the wave equation [8], [9] is challenging and often computationally prohibitive [10]. Recent work has shown promise that computational burden in seismogram inversion can be alleviated by using Principal Component Analysis (PCA) [11] for dimensionality reduction [12]. However, assuming linear correlations between parameters is often not justifiable, particularly for seismic data that typically exhibits strong nonlinear correlations. Thus, there is a great need for methods that will perform classification of seismic data in a computationally efficient way while preserving accuracy.

Classification of data through pattern recognition was first introduced nearly a century ago by Fisher [13]. Since then it has been used in a wide variety of disciplines including medical imaging [14], [15], subsurface hydrology [16], and remote sensing [17]. Within the context of seismic source classification problems, spectral methods [18], machine learning algorithms [19] and discriminant methods [20], [21] have been widely used. More recently, supervised learning algorithms, such as the Support Vector Machine (SVM) [22], [23], [24], [25], have found significant use in two and multi-level classification problems [26], [17]. In recently published research, SVMs were used to classify seismic data into earthquake and non-earthquake events [27].

Formally, SVMs determine a hyperplane to partition the data into disjoint sets. However, defining this hyperplane is often impossible in seismic classification problems because the relation between acoustic subsurface properties and seismic data is highly nonlinear. In order to remedy this issue, we employ kernel Principal Component Analysis (kPCA) [28], [29], [30], [31]. Similar to nonlinear SVM, kPCA maps the data to a higher dimensional feature space where linear PCA [11] is used to reduce the dimension of the data in feature space.

In this paper, we adopt the combination of nonlinear SVM with kPCA for classifying seismic data, as has been done recently in other fields like facial recognition [32] and medical (ECG) data [33]. The article is organized as follows. We begin by briefly recounting the mathematical formulation of the techniques, SVM and kPCA in Section II. Next, we describe the simulations used to generate our synthetic data, and define performance metrics in Section III. We present the results in Section IV. We conclude with a discussion of the key results in Section V.

II. Method

Our goal is to characterize how combining nonlinear SVM with kPCA performs in terms of computational expense, accuracy, and robustness when compared to nonlinear SVM alone. This combined method should be useful in the context of seismic data classification, where large data sets and the high nonlinear correlations necessitate nonlinear dimension reduction techniques. Both linear SVM and PCA perform poorly on the data in the sample space due to intrinsic nonlinearities and the results are not included. We describe the kPCA method briefly here for the sake of completeness

A. *Kernel principal component analysis*

Kernel component analysis (kPCA) is a nonlinear version of principal component analysis (PCA) [11], which is one of the most widely used dimensionality reduction techniques due to its conceptual and analytical simplicity and relative ease of implementation. However, appropriate application of PCA to a data set requires that the data has an orthogonal structure that can be used for the decomposition. Since seismic data is nonlinearly correlated, PCA is not useful in most cases. kPCA performs PCA based dimension-reduction on the data after a nonlinear transformation maps the data into a higher dimensional feature space $\mathcal{F}$, $\phi(\mathbf{x}) : \mathbb{R}^D \to \mathcal{F}$. The hope is that through the mapping $\phi(\cdot)$ features in the data are linearized and PCA can identify an orthogonal basis in feature space.

In practice however, kPCA does not use the mapping $\phi(\mathbf{x})$. Rather a kernel $\kappa(\cdot, \cdot)$ that satisfies Mercer's condition and is equivalent to $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ is used to construct a kernel matrix on which a singular value decomposition can be performed to identify the principal components of the data in $\mathcal{F}$. The effective application of kPCA hinges on selecting an appropriate κ that linearizes features in the data when mapped into $\mathcal{F}$. In this paper we use a Gaussian kernel

$$\kappa(\mathbf{x}, \mathbf{y}) = \exp\left[-\|\mathbf{x} - \mathbf{y}\|^2/(2\sigma^2)\right] \ . \tag{II.1}$$

which assigns greater weight to samples that are close to one another; close being defined in terms of the kernel parameter σ.

B. Dimensionality Reduction for Support Vector Machines

Support vector machines (SVM) use a maximal margin classifier to perform binary classification of a data set. Using training data described by a set of features, the method identifies boundary limits for each class in the feature space. These boundary limits are the local classifiers and the distance between the local classifiers is called the margin. SVM attempts to maximize this margin to identify the classifier as a hyperplane in the middle that separates the data into two groups while limiting the number of misclassifications. The data points on the boundaries support the limits and define the shape of the maximal margin classifier and are called the support vectors accordingly. The accuracy of the method is quantified by the percentage of samples that are correctly classified. However, in most cases the data may only be close to linearly separable. To account for this possibility we introduce slack variables that allow for misclassification and are controlled by a regularization factor, here denotes C.

The standard nonlinear extension of the SVM is to replace the linear norm with a kernel function. Note that the kernel function is the same as the kernel in the kPCA implementation. This leads to the SVM classification being controlled by

$$\kappa(\mathbf{a}, \mathbf{y}_i) + b = \langle \phi(\mathbf{a}), \phi(\mathbf{y}_i) \rangle + b \tag{II.2}$$

and the nonlinear SVM typically is applied to the entire transformed data without any dimension reduction. The disadvantage of this approach is that all features, equaling the number of sample points, are used in the classification to determine the vector $\phi(\mathbf{a})$. In kPCA-SVM the dimensionality of the space $\mathcal{F}$ is reduced first using kPCA and then the optimization problem to define the hyperplane is solved.

III. Classification of Reflectors

In order to characterize the utility of kPCA-SVM in seismic classification problems we apply the methodology to a synthetic data set of seismograms generated using the scalar wave equation with heterogeneous acoustic profile and perfectly matched boundary layer. Within each sample there is either a single reflector or two reflectors and we seek to partition the data into disjoint sets based on the number of reflectors. Seismograms are sampled at known locations and we vary the depths and magnitudes of the reflectors to increase variably in the data. The data is preprocessed by band-limiting the spectral magnitude of the wave form at each receiver. We perform kPCA-SVM according to the algorithm described in the previous section using the kPCA and SVM modules in the Python machine learning package Scikit-learn [34].

A. Model Setup

Synthetic seismograms are generated using a scalar wave equation with heterogeneous acoustic profile,

$$\partial_t^2 u(x,t) - c^2(x)\Delta u(x,t) = \omega(t)\delta(x - s_0) \tag{III.1}$$

with initial conditions

$$u(x,0) = 0, \quad \partial_t u(x,0) = 0 \ , \tag{III.2}$$

and first order perfectly matched layers along the boundaries of the domain. The term s_0 in (III.1) specifies the source position. Equation (III.1) is solved within the domain $D = [0.1, 1] \times [0.1, 0.8]$ for $t \in [0, 3.0]$. For $x = (x_1, x_2) \in D$, x_1 is the horizontal coordinate and x_2 is the vertical depth coordinate. Equation (III.1) is discretized using a second order accurate finite difference scheme and a first order non-reflecting boundary condition using the open source PySIT Python package [35].

Ninety equally spaced receivers are placed at depth $x_2 = 0.188$. We refer to a receiver position in D by r^i, $i = 1, 2, \ldots, 90$. The source position is fixed at $s_0 = (0.9, 0.188)$ and the source profile is modeled as a Ricker wavelet,

$$\omega(t) = (1 - 2\pi^2\nu_0^2(t - t_0)^2)e^{-\pi^2\nu_0^2(t-t_0)^2} \tag{III.3}$$

where $t_0 = \frac{6}{\pi\nu_0\sqrt{2}}$. The source's peak frequency is fixed at $\nu_0 = 10\,\mathrm{Hz}$ and the form for the offset, t_0, is chosen so that $\omega(t) = 0$ for $t < 0$.

The domain contains either one or two reflectors and is homogeneous otherwise. The depth profile of the horizontal reflectors is the derivative of a Gaussian function,

$$h(x_2; a, d) = \frac{a(x_2 - d)}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_2-d)^2} \ . \qquad \text{(III.4)}$$

The scaling parameter a determines the hardness of the reflector while d determines the reflector's depth. We call (a, d) a *scaling-depth pair*. Scaling-depth pairs for the horizontal reflector in set R_1 are randomly selected from $(a_1, d_1) \sim U([1.0, 5.0] \times [0.36, 0.76])$. Scaling-depth pairs for the upper reflector in the set R_2 are randomly selected from $(a_1, d_1) \sim U([1.0, 5.0] \times [0.36, 0.52])$ and scaling-depth pairs for the lower reflector in the set R_2 are randomly selected from $(a_2, d_2) \sim U([1.0, 5.0] \times [0.6, 0.76])$.

Receivers are placed at constant depth above the reflectors. The source position is fixed and the profile is modeled as a Ricker wavelet with peak frequency $10Hz$. The depth profile of the horizontal reflectors is the derivative of a Gaussian function. We select a constant background acoustic velocity of $c_0 = 2.0$. At each receiver, the acoustic pressure of each seismogram is Fourier transformed and band-limited prior to applying the kPCA-SVM classifier. This technique removes temporal dependence in the data and allows us to focus on similarities within the frequency profiles, which are expected to be a more telling signature of the number of reflectors in the system. Preliminary analysis (not included) revealed that wave interaction through multiple reflections was greater in the two-reflector seismograms than the single reflector case. This wave interaction was characterized by a multi-modal frequency spectrum. The single reflector system did not display such attributes. We down-sample the receiver signal and clip the late arrival time signal because it only contains very low amplitude reflections. Denoting a receiver signal at the i^{th} position as $r^i(t)$ the down-sampled and clipped discrete signal can be written as the vector

$$\mathbf{r}^i = (r^i(t_0), r^i(t_1), \ldots, r^i(t_{749}))^T, \qquad \text{(III.5)}$$

with $t_j = j\Delta t$, $\Delta t = \frac{1.25\,\mathrm{s}}{750}$. Next, we transform $\mathbf{r}^i$ to the frequency domain using a Fourier transform, bandlimit the frequency from 5–30 Hz using a hard cutoff, and collect the magnitude of the result.

Figure 1 shows the two seismogram profiles generated using the procedure described above. In both Figure 1 (a) and (b) the record is only shown for $t \in [0, 1.0]$ because the signal decays rapidly for $t > 1.0$. The source only seismogram is not shown in the figures in order to highlight the pressure signal generated from the reflectors. Our goal in applying kPCA-SVM is to identify whether a seismic signal is the result of one or multiple buried reflectors without resorting to seismic imaging methods.

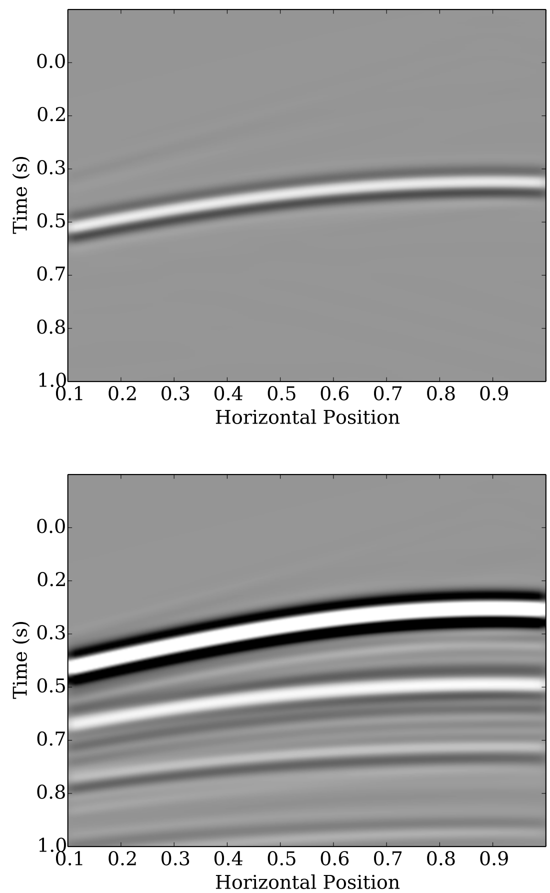

Fig. 1. (a) A sample seismogram generated using a single reflector. (b) A sample seismogram generated using two reflectors. In both (a) and (b) the record is only shown for $t \in [0, 1.0]$ because the signal decays rapidly for $t > 1.0$.

B. Parameter Exploration

The control parameters in these numerical experiments are: the number of dimensions that the data is projected onto in feature space m (the kPCA features), the Gaussian kernel parameter σ in (II.1), and the SVM regularization parameter C. We also investigate the influence of the number of receivers on the performance of our trained SVM.

The results of our parameter space exploration are presented in Table I. We consider data collected from one, two and three receivers for a data set composed of one-thousand samples, five hundred with one reflector and five hundred with two reflectors. The results of classification show that there was no significant change in the classification score (accuracy) with an increased number of receivers. Since kPCA only relies on the choice of the kernel, effective application hinges on selecting a suitable kernel that exposes features in the data when mapped into the higher dimensional feature space. We varied the Gaussian kernel parameter as follows, $\sigma = 0.5,\ 0.75,\ 1.0,\ 2.0$ as well as the margin parameters which relates to the number of points we allow to be mis-classified, $\log(C) = 4,\ 5,\ 6$. We also very the number or receivers used in the classification from one to three. In general, the maximum classification score is insensitive to σ.

TABLE I
KPCA-SVM OPTIMAL PERFORMANCE FOR THE THREE RECEIVER SCENARIOS. $n_f(\%)$: NUMBER OF KPCA FEATURES (PERCENT OF SPECTRUM); σ: KERNEL PARAMETER; C: SVM MARGIN PARAMETER; CV_{MAX}: MAXIMUM CV SCORE; CPU: CPU TIME FOR SVM FIT

Receivers	$n_f(\%)$	σ	$\log(C)$	CV_{max}
One	100(97.73)	0.5	5	0.933
One	110(99.57)	0.75	6	0.935
One	100(99.8)	1.0	6	0.935
One	150(100.0)	2.0	6	0.932
Two	200(97.96)	0.5	4	0.933
Two	320(99.94)	0.75	4	0.936
Two	240(99.92)	1.0	4	0.94
Two	150(99.96)	2.0	5	0.937
Three	480(99.96)	0.5	4	0.93
Three	210(98.98)	0.75	4	0.931
Three	90(97.52)	1.0	4	0.937
Three	210(99.97)	2.0	5	0.933

Preliminary numerical experiments showed that the standard deviation of the kernel σ did not have a significant effect on the accuracy of the SVM classifier. We select the optimal parameters obtained in those preliminary studies for our investigation of classification accuracy and CPU time and focus on the how classification accuracy and CPU time depends on the number of dimensions m that the data is projected onto in feature space and the regularization parameter C.

IV. KPCA-SVM COMPARISON

The main goals for our application of kPCA-SVM are to determine how dimensionality reduction influences the ability of an SVM to properly classify and distinguish between members of the two sets, and to quantify any efficiencies gained as a result. To this end, we systematically vary the number of kPCA features that are retained prior to classifying the data and the regularization parameter C and record both the classification accuracy and CPU time required. We use data collected at a single receiver for classification purposes.

Performance of the method is quantified in terms of (i) the average classification score on a given set of sample points as the fraction of the set classified correctly, and (ii) the CPU time required for optimization convergence. The average score is a reliable metric for two-level classification algorithms when the training and test sets are made up of equal samples from each class. Half of the samples are generated using a single reflector, and the other half are are generated using two reflectors. Stratified-shuffle-split cross-validation is used to form subsets of samples, 15% of the data, with half from each set. The kPCA-SVM scheme is trained on the remaining samples and the accuracy score is evaluated on the test samples. This process is repeated 100 times for each kPCA-SVM instance in order to perform cross-validation. This allows for reliable characterization of each kPCA-SVM instance's accuracy and robustness.

Figure 2 plots the contribution of the singular values to the partial sums normalized by the sum of the entire spectrum against the singular value index for a set composed of one thousand samples, five hundred with one reflector and five hundred with two reflectors. The graph shows that over 90% variation in the spectrum of the kernel matrix can be accounted for by the first ten singular values and that the first one hundred singular values account for over 99% of the spectrum. This suggests that kPCA-SVM should obtain near optimal performance when one hundred features are used and that no additional benefits are gained thereafter. We use the word optimal in terms of the highest value of the accuracy score averaged over the one hundred cross-validation instances.

A. *Classification Score*

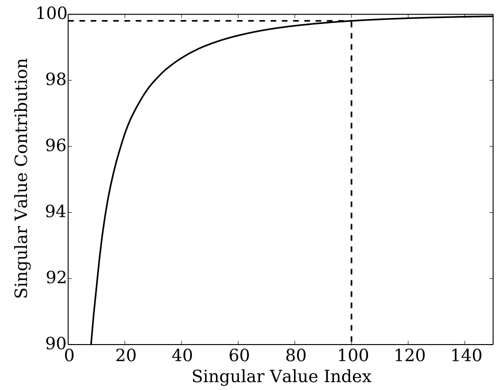

Fig. 2. The contribution of the singular values to the partial sums normalized by the sum of the entire spectrum for the first 150 singular values of the kernel matrix constructed using the thousand samples. At 100 features, (10% of the total number of features) the singular values nearly capture the total variation of the sample set in feature space. The dashed line indicates the optimal number of kPCA features to maximize the accuracy in the independent test set. The large amount of the spectrum contained in the first 100 singular values indicates that using the subspace defined by these singular values should provide sufficient information for proper classification.

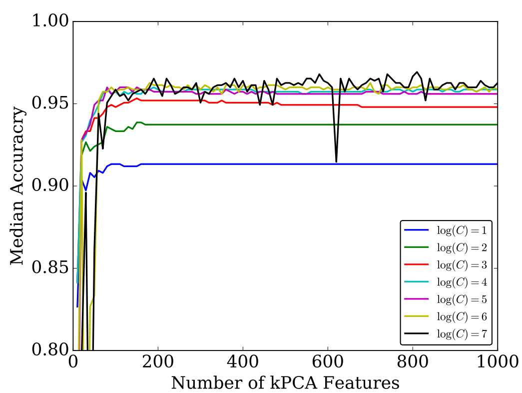

Fig. 3. Mean SVM cross-validated classification scores for the range of regularization parameters considered. The scores are computed over successive kPCA features considered in the SVM. Scores improve as the number of features considered increases up to one hundred features and then decreases with a plateau reached after 200 kPCA features.

Our principal study uses a data set composed of five thousand samples, evenly split between one and two reflectors. We compute and compare the SVM cross-validated accuracy scores between three values for the margin parameter: $\log(C) = 1, 2, 3, 4, 5, 6, 7$ while changing the number of retained kPCA features from ten to five thousand. When all five thousand features are used, kPCA-SVM is standard nonlinear SVM. Figure 3 shows the classification scores for the one hundred cross validation tests for each of the values of C considered (increasing in value from left to right) plotted as a function of number of features retained. Only values of one to one-thousand are shown as the values beyond one-thousand are relatively constant. All value tests reach their optimal performance after approximately two hundred features are used ($\approx 5\%$ of the total number of features). Lower values of C result in lower accuracy, while higher values result in better performance. The values seems to converge around 95% accuracy and no additional accuracy is gained by further increasing C. However, the highest values of C, 10^6 and 10^7, are less stable in their accuracy. This is likely due to the higher values corresponding to less regularization and thus more difficult optimization problems.

B. *CPU Times*

Figure 4 shows the mean CPU time required to solve the kPCA-SVM classification problem plotted against the number of features used in classification for the single receiver case with margin parameter values of $\log(C) = 1, 2, 3, 4, 5, 6, 7$. In general, larger values of the margin parameters results in slower CPU times because the optimization problem is harder to solve due to less regularization. Conversely, smaller values of C result in less CPU time, at the expense of accuracy, cf. Fig. 3. In all values considered, CPU time increases significantly with the number of features used. In some cases, using a lower number of features results in up to five times speed up when compared to standard nonlinear SVM.

Figure 5 plots the median accuracy score against the median CPU time for all values of C considered when 100 features (circles) and 1000 features (squares) are used. This plot combines information from Fig. 3 and Fig. 4. At low values of C the accuracy is lower, but so is the CPU time. As values of C increase, so does the accuracy but so does the CPU time. However, after $\log(C) = 4$ no improvement of accuracy is observed, but the CPU time still increases. This behavior is observed for both 100 or 1000 features. No significant increase in accuracy is observed using 1000 features when compared to 100 features. However there is a considerable

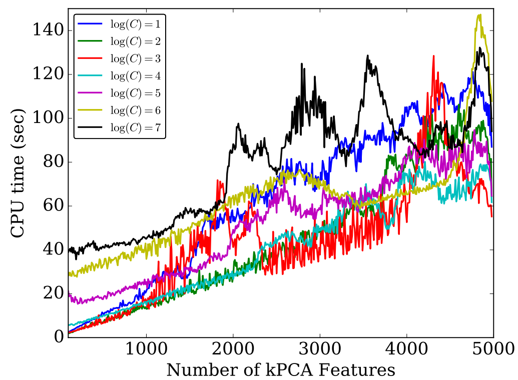

Fig. 4. CPU time for SVM training vs. number of kPCA features used. CPU times are shown for the single receiver case.

increase in required CPU time. This difference is larger for smaller values of C than for larger values. In general the two parameters, compete with one another; smaller values of C result in lower CPU times at the expense of accuracy.

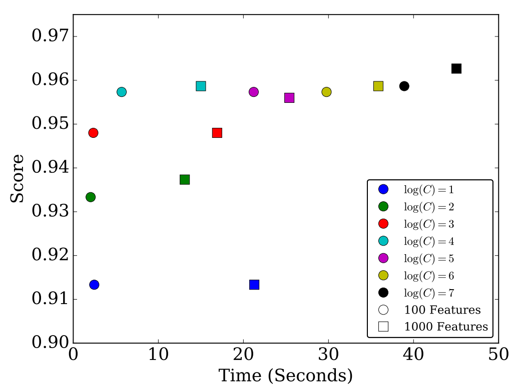

Fig. 5. Median accuracy score plotted as a function of the median CPU time for all values of C considered when 100 features (circles) and 1000 features (squares) are used. At low values of C the accuracy is lower but so is the CPU time. As values of C increase, so does the accuracy but so does the CPU time.

V. Discussion

Although nonlinear SVMs have been previously used in seismic problems [27], we report the first investigation of how varying the dimension of the subspace on which a nonlinear SVM is trained in conjunction with the regularization parameter influences the classifier's robustness and computational speed. Specifically, the method advances the use of nonlinear SVM in seismic applications by quantifying the impact of performing dimensionality reduction based on kPCA on the robustness of classification obtained using SVM. We found that kPCA-SVMs provide a robust classification algorithm, obtaining the same accuracy as nonlinear SVM when a small fraction (less than 10%) of the total number of kPCA features are used with up to five times speedup.

We also found that both kPCA-SVM and nonlinear SVM classify more than 90% of the data correctly and the regularization parameter had more of an effect on accuracy than the number of features used once an optimal number of features was used. For large data sets, CPU time increases considerably when more kPCA features are included, but the SVM classifier score does not improve. This behavior suggests that for large problems, such as those encountered in real data, dimension reduction prior to classification could significantly reduce computation times without loss of accuracy.

The synthetic seismogram data analyzed in this study is representative of many real world seismic situations including the monitoring of faults, detection of underground nuclear weapon testing, and surveillance of facility usage patterns. In the field, seismic inversion relies on assuming that the seismograms given were generated by a particular model of seismic wave propagation and a particular subset of acoustic profiles. The presented method of kPCA-SVM could be applied to determine which seismic wave propagation model is best suited to invert a particular seismogram. The development of these type of classification techniques is needed in situations where limited seismic observation data is available and characterizing the source or acoustic profile model is a key first step in full seismic inversion. Further developments of the method will include: incorporating a heterogeneous medium, using multi-level SVM classification to discover the number reflectors present, and developing a methodology for learning the optimal kernel for feature extraction.

Acknowledgments

The authors would like to thank the DOE-LDRD program for supporting this work. JDH thanks LANL's DR research program, Grant # 20140002DR, the support of the Center for Nonlinear Studies at LANL, Grant # DE-AC52-

06NA25396, the Los Alamos Director's Postdoctorial Fellowship 20150763PRD4.

References

[1] E. L. Majer, R. Baria, M. Stark, S. Oates, J. Bommer, B. Smith, and H. Asanuma, "Induced seismicity associated with enhanced geothermal systems," *Geothermics*, vol. 36, no. 3, pp. 185–222, 2007.

[2] H. Sato, M. C. Fehler, and T. Maeda, *Seismic wave propagation and scattering in the heterogeneous earth*. Springer, 2012, vol. 484.

[3] C. Bradley and E. Jones, "Modeling propagation effects from explosions in western china and india," Los Alamos National Lab., NM (United States), Tech. Rep., 1998.

[4] M. Denny and J. Zucca, "Introduction: Doe non-proliferation experiment," *Arms Control and Nonproliferation Technologies, DOE/AN/ACNT-94A*, pp. 8–21, 1994.

[5] B. W. Stump, "Practical observations of us mining practices and implications for ctbt monitoring," Los Alamos National Lab., NM (United States), Tech. Rep., 1995.

[6] W. L. Ellsworth, "Injection-induced earthquakes," *Science*, vol. 341, no. 6142, p. 1225942, 2013.

[7] R. Vidic, S. Brantley, J. Vandenbossche, D. Yoxtheimer, and J. Abad, "Impact of shale gas development on regional water quality," *Science*, vol. 340, no. 6134, p. 1235009, 2013.

[8] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler, "A computational framework for infinite-dimensional bayesian inverse problems part i: The linearized case, with application to global seismic inversion," *SIAM J Sci Comput*, vol. 35, no. 6, pp. A2494–A2523, 2013.

[9] J. Martin, L. C. Wilcox, C. Burstedde, and O. Ghattas, "A stochastic newton mcmc method for large-scale statistical inverse problems with application to seismic inversion," *SIAM J Sci Comput*, vol. 34, no. 3, pp. A1460–A1487, 2012.

[10] M. Fehler, "Seam update: Seam phase i rpsea update: Status of simulations," *The Leading Edge*, vol. 31, no. 12, pp. 1424–1426, 2012.

[11] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.

[12] J. Rezaie, J. Saetrom, E. Smorgrav *et al.*, "Reducing the dimensionality of geophysical data in conjunction with seismic history matching," in *SPE Europec/EAGE Annual Conference*. Society of Petroleum Engineers, 2012.

[13] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann Eugenic*, vol. 7, no. 2, pp. 179–188, 1936.

[14] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE T Med Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.

[15] A. Papadopoulos, D. I. Fotiadis, and A. Likas, "Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines," *Artif Intell Med*, vol. 34, no. 2, pp. 141–150, 2005.

[16] B. Wohlberg, D. M. Tartakovsky, and A. Guadagnini, "Subsurface characterization with support vector machines," *IEEE T Geosci Remote*, vol. 44, no. 1, pp. 47–57, 2006.

[17] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J Photogramm*, vol. 66, no. 3, pp. 247–259, 2011.

[18] B. P. Allmann, P. M. Shearer, and E. Hauksson, "Spectral discrimination between quarry blasts and earthquakes in southern california," *Bull. Seismol. Soc. Am.*, vol. 4, no. 98, pp. 2073–2079, 2008.

[19] T. Tiira, "Discrimination of nuclear explosions and earthquakes from teleseismic distances with a local network of short period seismic stations using artificial neural networks," *Phys. Earth Planet Inter.*, vol. 1–4, no. 97, pp. 247–2689, 1996.

[20] I. Che, M. Jun, and J. Jeaon, "A compound linear discriminant method for small magnitude seismic events and its application to the north korea seismic event of october 9, 2006," *Earth, Planets, Space*, vol. 12, no. 59, pp. e41–e44, 2007.

[21] H. Kuyuk, E. Yildirim, E. Dogan, and G. Horasan, "Clustering seismic activities using linear and nonlinear discriminant analysis," *J. Earth Sci.*, vol. 1, no. 25, pp. 140–145, 2014.

[22] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[23] M. D. Gonzalez-Lima, W. W. Hager, and H. Zhang, "An affine-scaling interior-point method for continuous knapsack constraints with application to support vector machines," *SIAM J Optimiz*, vol. 21, no. 1, pp. 361–390, 2011.

[24] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.

[25] B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[26] M. C. Ferris and T. S. Munson, "Interior-point methods for massive support vector machines," *SIAM Journal on Optimization*, vol. 13, no. 3, pp. 783–804, 2002.

[27] J. Kortstrom, M. Uski, and T. Tiira, "Automatic classification of seismic events within a regional seismograph network," *Comput Geosci*, vol. 87, no. 2, pp. 22–30, 2016.

[28] K. I. Kim, K. Jung, and H. J. Kim, "Face recognition using kernel principal component analysis," *IEEE Signal Proc Let*, vol. 9, no. 2, pp. 40–42, 2002.

[29] B. Schölkopf, S. Mika, C. J. Burges, P. Knirsch, K.-R. Müller, G. Rätsch, and A. J. Smola, "Input space versus feature space in kernel-based methods," *IEEE T Neural Networ*, vol. 10, no. 5, pp. 1000–1017, 1999.

[30] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Artificial Neural Networks—ICANN'97*. Springer, 1997, pp. 583–588.

[31] ——, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput*, vol. 10, no. 5, pp. 1299–1319, 1998.

[32] W. Li, W. Gong, Y. Liang, and W. Chen, "Feature selection based on kpca, svm and gsfs for face recognition," in *Pattern Recognition and Image Analysis*. Springer, 2005, pp. 344–350.

[33] L. Kanaan, D. Merheb, M. Kallas, C. Francis, H. Amoud, and P. Honeine, "Pca and kpca of ecg signals with binary svm classification," in *Signal Processing Systems (SiPS), 2011 IEEE Workshop on*. IEEE, 2011, pp. 344–348.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[35] R. Hewett, L. Demanet, and the PySIT Team, "PySIT: Python seismic imaging toolbox v0.5," 2013, release 0.5. [Online]. Available: http://www.pysit.org